

# Automatización de un registro hospitalario de tumores

Josepa Ribes<sup>a</sup> / Jordi Gálvez<sup>a</sup> / Àngels Melià<sup>a</sup> / Ramon Clèries<sup>a</sup> / Xavier Messeguer<sup>b</sup> / Francesc Xavier Bosch<sup>a</sup>

<sup>a</sup>Servei d'Epidemiologia i Registre del Càncer. Institut Català d'Oncologia. L'Hospitalet de Llobregat. Barcelona. España.

<sup>b</sup>Departament de Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya. Barcelona. España.

(Automatization of a hospital-based tumor registry)

## Resumen

**Introducción:** El Instituto Catalán de Oncología automatizó los procedimientos manuales de captación de la información de las bases de datos del alta hospitalaria (AH) y anatomía patológica (APA) mediante una aplicación informática (ASEDAT) con el objetivo de aumentar la fiabilidad de los datos y reducir los costes del Registro Hospitalario de Tumores (RHT).

**Material y método:** ASEDAT detecta los tumores incidentes del centro a partir de las bases de datos de APA y de las AH mediante la selección de la información básica para cada uno de ellos. Se resolvió el RHT para el período 1999-2000 mediante el procedimiento manual y automatizado, y se compararon entre sí los resultados.

**Resultados:** Se detectaron 10.498 pacientes oncológicos. La resolución manual detectó 8.309 tumores incidentes y 2.374 tumores prevalentes. ASEDAT resolvió automáticamente 8.901 pacientes (84,8%), en los cuales se detectaron 8.367 tumores incidentes, 58 tumores más que con el procedimiento manual. La validación de la concordancia se realizó en los tumores incidentes detectados por ambos métodos (7.063 tumores). En 6.185 tumores (87,6%), la información coincidió en todas las variables. De los tumores discordantes, 692 (9,8%) fueron generados por el personal del RHT en la resolución manual y el resto (n = 186; 2,6%) por la aplicación (resolución automática).

**Conclusiones:** La automatización de un registro de cáncer es posible siempre y cuando el centro disponga de las bases de datos de APA y AH codificadas e informatizadas.

**Palabras clave:** Cáncer. Registro. Automatización.

## Abstract

**Introduction:** To increase data reliability and reduce the costs associated with the HTR, the Catalan Institute of Oncology programmed the manual procedures of data collection from databases by means of a computer application (ASEDAT).

**Material and method:** ASEDAT detects the incident tumors of the registry from the databases of the pathology records (PR) and discharge records (DR) and selects the basic information from both databases. Data from the HTR data was collected for the period 1999-2000 by means of 2 procedures: manual and automatized collection and the results obtained were compared.

**Results:** 10,498 cancer patients were detected. Manual resolution detected 8,309 incident tumors and 2,374 prevalent tumors. ASEDAT automatically detected 8,901 patients (84.8%), in whom 8,367 incident tumors were detected (58 more tumors than the manual procedure). Validation of agreement was performed in the incident tumors detected by both methods (7,063 tumors). In 6,185 tumors (87.6%) the information agreed in all the variables. Of the discordant tumors, 692 (9.8%) were obtained by the RHT staff using manual resolution, and the remainder (186; 2.6%) were obtained by the application (automatic resolution).

**Conclusions:** Cancer registry automatization is feasible when PR and DR databases are available, coded and automatized.

**Key words:** Cancer. Registry. Automatization.

## Introducción

El cáncer es uno de los problemas sanitarios más importantes en Cataluña y se estima que en el año 2000 se diagnosticaron unos 28.000 nuevos casos de cáncer, de los cuales fallecieron

15.000. La incidencia del cáncer en Cataluña muestra una tendencia creciente desde 1980 y las proyecciones de la incidencia para el año 2005 sugieren un incremento del 20% en varones y del 14% en las mujeres<sup>1</sup>. Un elemento básico para el control del cáncer en un país es la disponibilidad de sistemas de información que elaboren datos disponibles de cáncer con el fin de evaluar las necesidades asistenciales de los pacientes oncológicos, la utilización de los servicios sanitarios y los progresos en el control de la enfermedad. En Cataluña, el Plan Director de Oncología recomienda la creación del Registro de Cáncer de Catalunya (RCC) para poder recoger información de esta enfermedad en todos los ámbitos territoriales<sup>2</sup>. En Cataluña, además de una larga tradición de registros de cáncer hospitalarios<sup>3</sup>, se dispone de 2 registros de cáncer poblacionales (RCP):

Correspondencia: Josepa Ribes Puig.

Servei d'Epidemiologia i Registre del Càncer.

Institut Català d'Oncologia.

Gran Via, s/n, km 2,7. 08907 L'Hospitalet de Llobregat.

Barcelona. España.

Correo electrónico: j.ribes@iconcologia.net

Recibido: 22 de julio de 2004.

Aceptado: 20 de diciembre de 2004.

el de Tarragona, que recoge datos desde 1980<sup>4</sup>, y el de Girona, que recoge datos desde 1994<sup>4,5</sup>.

El Instituto Catalán de Oncología (ICO) y el Hospital Universitario de Bellvitge (HUB), situados en L'Hospitalet de Llobregat (Barcelona), comparten el Registro Hospitalario de Tumores (RHT) que se inició en octubre de 1990. El RHT tiene registrados hasta el año 2003 unos 50.000 tumores malignos correspondientes a unos 3.700 tumores incidentes al año. Las fuentes de información básicas del RHT son: el servicio de anatomía patológica (APA) y las altas hospitalarias (AH). Anatomía patológica dispone de una base de datos informatizada desde 1988 con los diagnósticos de las biopsias, citologías y necroscopias, codificadas según una adaptación de la normativa de la Universidad Católica de Lovaina<sup>6</sup>. Esta clasificación se basa en codificar la topografía, la histología y el comportamiento de los tumores de forma similar a la realizada por la clasificación SNOMED<sup>7</sup>. El servicio de Documentación Médica dispone de una base de datos informatizada también desde 1988 que registra todos los diagnósticos y procedimientos de las AH, codificadas según las normas de la novena edición de la Clasificación Internacional de Enfermedades (CIE-9)<sup>8</sup> y las recomendaciones del Conjunto Mínimo Básico de Datos (CMBD)<sup>9</sup>. El RHT codifica los tumores según las normas de la International Classification of Diseases for Oncology (ICD-O, primera edición para la localización tumoral<sup>10</sup> y segunda edición para la histología<sup>11</sup>).

Para la elaboración manual del RHT en el período 1990-2000 se fusionaban informáticamente las bases de datos de AH, APA y RHT para extraer, de todos los pacientes diagnosticados de cáncer en un año concreto, toda la información sanitaria actual y retrospectiva. Las 3 bases de datos se conectaban mediante el número de historia clínica. Con esta información se elaboraba anualmente un listado que contenía los datos de cerca de 5.000 pacientes, entre los que había tumores incidentes y prevalentes. Este listado era revisado por un experto, que extraía para cada tumor la localización, la morfología, la fecha de diagnóstico, la base diagnóstica y los servicios hospitalarios que han diagnosticado y/o tratado el tumor según las normas de la European Network of Cancer Registries (ENCR)<sup>12</sup> y la International Agency for Research on Cancer (IARC)<sup>13</sup>. La información seleccionada se introducía manualmente en la base de datos del RHT después de recodificarse a la ICD-O. La información de los tumores prevalentes del RHT se contrastaba anualmente con la nueva información aportada por APA y AH con el fin de actualizarla. Los análisis estadísticos descriptivos y la tabulación de resultados se realizaban por el personal del RHT y la publicación se encargaba a una empresa externa<sup>14</sup>.

Este proceso presentaba algunos inconvenientes: a) el coste de tiempo para proporcionar los datos anua-

les (2 meses para la extracción de información del listado y su recodificación, 1 mes para la introducción de los datos y 3 meses para la validación, los análisis estadísticos, la tabulación de resultados y su publicación); b) el coste económico (personal médico, estadístico, administrativo y contratación de la imprenta para la publicación); c) los errores en los procesos de resolución, extracción, recodificación e introducción de la información, y d) el volumen de los casos registrados que, al aumentar anualmente, incrementa el coste del RHT. Según la experiencia de algunos registros de cáncer poblacionales automatizados<sup>15</sup>, se propuso automatizar los procedimientos manuales descritos del RHT del ICO y el HUB con el objetivo de reducir los costes y el porcentaje de errores.

## Material y método

A partir del listado resultante de la fusión de los datos de AH y APA, se diseñó una aplicación informática para obtener los casos incidentes del RHT automáticamente, reproduciendo las normas especificadas por la IARC<sup>15</sup>. La aplicación, denominada ASEDAT (Algoritmo de Selección y Extracción de DATos Tumoraes), incluyó diversas etapas en el procesamiento de la información. La primera etapa fue la conversión de las clasificaciones de Lovaina de APA y de la CIE-9 de las AH a la recomendada para los registros de cáncer, la clasificación ICD-O. Los códigos morfológicos de APA se recodificaron a la clasificación de la ICD-O, tercera edición<sup>16</sup>, y los códigos topográficos de APA y de las AH a la ICD-O, primera edición<sup>10</sup>. Diversos códigos topográficos de la clasificación ICD-O pueden referirse a una misma localización tumoral (códigos equivalentes de localización), asignándolos a un grupo de equivalencia topográfico según las normas propuestas por el Ministerio de Salud de Dinamarca<sup>15</sup>. Por ejemplo, los códigos topográficos incluidos en el rango 162,0 a 162,9 corresponden a la localización pulmonar (códigos equivalentes) y pertenecen al grupo de equivalencia topográfico 20. Asimismo, los códigos morfológicos se agruparon según su comportamiento clínico. Por ejemplo, los códigos morfológicos en los que los 3 primeros dígitos son 817 o 818 y el comportamiento es 3, 6 o 9 corresponden al carcinoma hepatocelular (códigos equivalentes) y al grupo de equivalencia morfológico 10. La definición de estos códigos equivalentes y de sus grupos correspondientes permitió diferenciar los tumores según el grado de dificultad de resolución: a) fáciles, cuando las localizaciones de AH y de APA coinciden (los códigos topográficos son equivalentes y por tanto pertenecen a un mismo grupo), y b) difíciles, cuando dos o más localizaciones no coinciden (los códigos topográficos no son equivalentes y no pertenecen a un

mismo grupo). En este último caso, la aplicación decide si las localizaciones proceden del mismo tumor (tumor primario con metástasis) o de tumores primarios distintos (tumores múltiples).

La resolución de los tumores difíciles requiere la elaboración de un algoritmo informático complejo que incluya criterios de proximidad anatómica de las localizaciones tumorales, equivalencia de las morfologías asociadas con cada localización tumoral y frecuencia de localización de las metástasis según el órgano en el que se originó el tumor. Si 2 localizaciones tumorales no están próximas anatómicamente y las morfologías difieren (códigos morfológicos no equivalentes), se consideran tumores múltiples. En esta situación, se registrará la información correspondiente a cada tumor. Los casos en que las localizaciones tumorales están anatómicamente próximas y las morfologías son equivalentes, el algoritmo considerará que se tratan del mismo tumor. En este caso, se escoge la codificación de la localización tumoral más específica. Los casos que no cumplen ninguna de estas 2 condiciones, el algoritmo considera que deben resolverse de forma manual por un experto<sup>17</sup>.

Una vez identificados los tumores candidatos a ser registrados en el RHT, la aplicación seleccionará automáticamente la información de las variables esenciales de cada tumor (tabla 1) a partir de la información de AH y APA. La morfología tumoral se selecciona según el grado de validez diagnóstica de APA. La aplicación escoge primero la morfología procedente de las biopsias realizadas en el tumor primario, en segundo lugar de las biopsias procedentes de metástasis y, por último, de las citologías. Si coexiste más de una morfología equivalente con el mismo grado de validez diagnóstica, la aplicación selecciona la más específica (código de mayor especificidad de la clasificación ICD-O). El método diagnóstico incluido en el RHT será el de mayor validez diagnóstica. Si no hay diagnósticos en APA, el método diagnóstico se considera desconocido. La fecha de incidencia del tumor en el centro sanitario es la fecha más antigua asociada al tumor (AH o APA)<sup>17</sup>. Si en la base de datos de APA consta el TNM patológico, éste se recopilará automáticamente en el RHT según la normativa de la International Union Against Cancer<sup>18</sup>. Los servicios clínicos que han generado información sobre el diagnóstico y/o el tratamiento del tumor se incluirán en el RHT.

El responsable del RHT deberá resolver los tumores incidentes que la aplicación selecciona para ser revisados manualmente: a) tumores de AH codificados como malignos, pero que en los diagnósticos de APA posteriores al alta no se confirma la malignidad; b) imposibilidad de discernir si un paciente presenta tumores múltiples o un único tumor, y c) los tumores hepáticos sin especificar si son primarios o metastásicos (código ICD-9: 1552). Los tumores prevalentes (los que

**Tabla 1. Variables esenciales del registro hospitalario de tumores**

VARIABLES ESENCIALES DEL RHT

Identificación personal: nombre, sexo, fecha de nacimiento, residencia habitual, DNE, CIP<sup>a</sup>  
 Fecha de incidencia<sup>b</sup>  
 Localización tumoral  
 Morfología/comportamiento  
 Base del diagnóstico<sup>c</sup>  
 Servicios médicos y quirúrgicos del centro sanitario  
 pTNM  
 Fuentes de información

<sup>a</sup>Información extraída del fichero de admisiones del centro hospitalario. <sup>b</sup>Fecha del primer evento del tumor en el centro sanitario. <sup>c</sup>Citología: examen de células procedentes del tumor primario o de metástasis, incluido el aspirado de fluidos mediante endoscopia o agujas examen de sangre periférica y aspirado de médula ósea. Histología de metástasis: examen histológico de tejido procedente de metástasis incluidas las muestras procedentes de necropsias. Histología del tumor primario: examen histológico procedente de tejido del tumor primario, incluidas las biopsias de médula ósea y las procedentes de necropsias. No microscópica: cuando no se evidencia en la base de datos de anatomía patológica ninguna de las opciones anteriores.

ya estaban incluidos en el RHT) deberán compararse manualmente con la información más reciente (generada por la aplicación) para comprobar si es candidata a ser actualizada.

Para validar la aplicación se procesaron los datos del período 1999-2000 del RHT del ICO-HUB de forma manual y automatizada (10.498 pacientes). Los casos se seleccionaron entre todos los pacientes asistidos en los 2 centros en el período de estudio que presentaron al menos un código de malignidad en la base de datos de APA (códigos morfológicos de Lovaina: 77 al 99) y/o en la base de datos de las AH (códigos de la ICD-9: 140,0 al 208,9). Se compararon la resolución y la información de los casos incidentes extraídos mediante los 2 métodos y se determinó el grado de concordancia a partir del índice de kappa<sup>19</sup>. La revisión de las historias clínicas de los casos discordantes permitió mejorar las reglas implementadas en la aplicación.

Para la difusión de los resultados del RHT del ICO/HUB se implementó en la página web del ICO un formulario para que los usuarios realicen consultas abiertas según condiciones específicas (localización tumoral, morfología, edad, sexo, año o base diagnóstica)<sup>20</sup>. Como resultado se obtiene la distribución de los tumores según las condiciones solicitadas.

Para evaluar la exportación de la aplicación a otros centros sanitarios, ésta se implementó en el Hospital de la Cruz Roja de l'Hospitalet de Llobregat, que disponía de un registro de tumores desde 1999. Se compararon los resultados obtenidos del RHT para el año 1999 mediante el procedimiento manual y el automatizado. La

Figura 1. Distribución de los tumores en incidentes y prevalentes según la aplicación ASEDAT.

| Revisar | Resultados | Resultados Originales | Filtros | HC     | Data | Loc             | Morf | DX Servs | Prevalent | Revisar |
|---------|------------|-----------------------|---------|--------|------|-----------------|------|----------|-----------|---------|
| 100511  |            | Revisar 1             |         |        |      |                 |      |          |           |         |
| 100591  |            | 04/05/2001            | 174.9   | M85002 | 7    | GIN CMM UMA     |      | APA+AH   |           |         |
| 100785  |            | 23/04/2001            | 188.9   | M81303 | 7    | URO             |      | APA+AH   |           |         |
| 101064  |            | 16/03/2001            | 195.9   | M81403 | 7    | URO             |      | APA+AH   |           |         |
| 101267  |            | 21/12/2000            | 188.0   | M83123 | 7    | URO COT         |      | 1 APA+AH |           |         |
| 101327  |            | 16/07/2001            | 193.9   | M80716 | 6    | RAD             |      | APA      |           |         |
| 101338  |            | 15/11/2000            | 196.9   | M95903 | 7    | NEF CIR HEM ORL |      | APA+AH   |           |         |
| 101429  |            | 06/06/2001            | 173.9   | M97003 | 7    | DER             |      | APA      |           |         |
| 101466  |            | 05/08/2001            | 188.9   | M98233 | 7    | HEM             |      | APA+AH   |           |         |
| 101958  |            | Revisar 1             |         |        |      |                 |      |          |           |         |
| 101963  |            | 15/03/2001            | 189.0   | M81203 | 7    | URO             |      | APA      |           |         |
| 101958  |            | 14/09/2001            | 185.9   | M80103 | 7    | URO             |      | APA      |           |         |
| 102741  |            | 03/07/2001            | 148.1   | M80703 | 7    | ORL             |      | APA+AH   |           |         |
| 102943  |            | Revisar 1             |         |        |      |                 |      |          |           |         |
| 103009  |            | 23/10/2000            | 196.9   | M95903 | 7    | HEM CIR         |      | 1 APA+AH |           |         |
| 103055  |            | 14/02/2001            | 156.1   | M80003 | 9    | CIR             |      | AH       |           |         |
| 103113  |            | 20/07/2001            | 174.0   | M85303 | 7    | CIR RAD UMA     |      | APA+AH   |           |         |
| 103152  |            | 10/05/2001            | 175.6   | M80713 | 7    | CIR             |      | APA      |           |         |
| 103321  |            | 06/07/2000            | 188.9   | M81303 | 7    | URO             |      | APA+AH   |           |         |
| 10391   |            | 23/07/1999            | 190.9   | M81403 | 7    | GIN MED         |      | 1 APA+AH |           |         |
| 104046  |            | 09/02/1999            | 169.9   | M96533 | 7    | NCR HEM ???     |      | 1 APA+AH |           |         |
| 104137  |            | Revisar 1             |         |        |      |                 |      |          |           |         |
| 104170  |            | 26/10/2001            | 153.2   | M81403 | 7    | CIR DNC DIG     |      | APA+AH   |           |         |
| 104237  |            | 09/11/1997            | 185.9   | M81403 | 7    | URO NEM ???     |      | 1 APA+AH |           |         |
| 104316  |            | NO 1                  |         |        |      |                 |      |          |           |         |
| 104710  |            | Revisar 1             |         |        |      |                 |      |          |           |         |
| 10481   |            | 28/05/2000            | 153.6   | M81403 | 7    | DIG CIR         |      | APA+AH   |           |         |
| 10483   |            | 11/01/2000            | 193.9   | M85103 | 7    | CIR             |      | APA+AH   |           |         |
| 109061  |            | 30/11/2001            | 153.2   | M81403 | 7    | CIR             |      | APA+AH   |           |         |
| 105134  |            | 11/10/2001            | 162.3   | M80713 | 7    | NEM MED         |      | APA+AH   |           |         |
| 105361  |            | 11/09/2001            | 166.9   | M80003 | 9    | URO             |      | AH       |           |         |
| 106100  |            | 06/02/1995            | 173.4   | M80713 | 7    | MAL ???         |      | 1 APA+AH |           |         |
| 106249  |            | 01/09/2001            | 174.9   | M80003 | 9    | MED             |      | AH       |           |         |
| 106759  |            | Revisar 1             |         |        |      |                 |      |          |           |         |
| 106833  |            | 27/09/2001            | 196.9   | M95903 | 9    | HEM             |      | AH       |           |         |
| 1069    |            | 05/09/2001            | 173.0   | M80703 | 7    | DER             |      | APA      |           |         |
| 106926  |            | Revisar 1             |         |        |      |                 |      |          |           |         |
| 106942  |            | Revisar 1             |         |        |      |                 |      |          |           |         |
| 10719   |            | Revisar 1             |         |        |      |                 |      |          |           |         |
| 107239  |            | Revisar 1             |         |        |      |                 |      |          |           |         |
| 107845  |            | 18/10/1988            | 174.8   | M85303 | 7    | GIN DNC ???     |      | APA+AH   |           |         |
| 10794   |            | 06/03/2001            | 174.9   | M85003 | 7    | CIR             |      | APA+AH   |           |         |

aplicación se modificó para aceptar la información de APA codificada según las normas del SNOMED<sup>7</sup>.

## Resultados

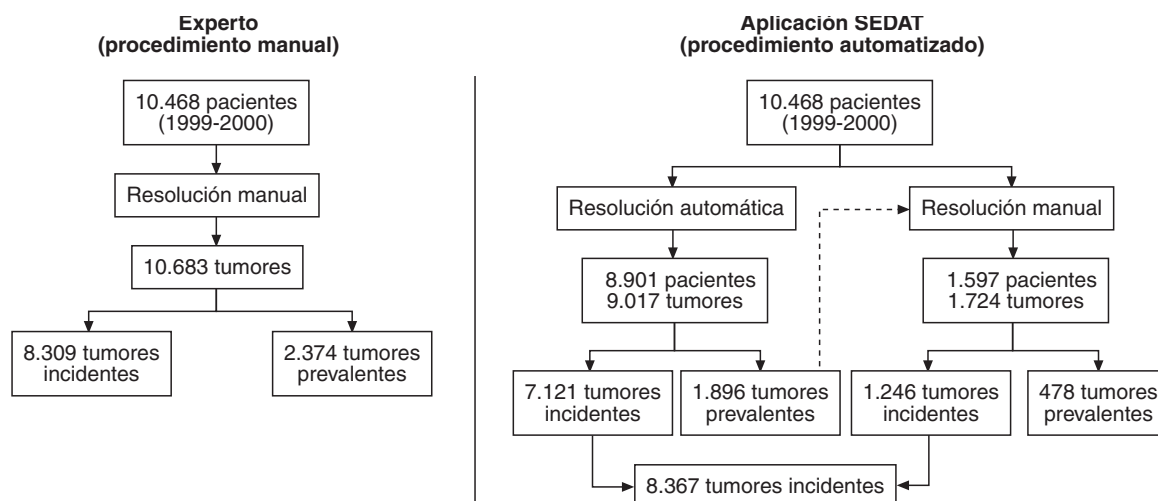
La aplicación ASEDAT excluye los tumores con fecha de incidencia anterior a la de inicio del RHT del ICO-HUB, quedando excluidos los tumores que ya habían sido asistidos en el centro antes del 1 de octubre de 1990. Otra condición de exclusión fueron los tumores basocelulares de la piel y los carcinomas *in situ*, que por motivos logísticos se excluyeron en el RHT desde su inicio. La aplicación diseñada permite escoger la opción de definir la fecha de inicio del RHT (que difiere en cada centro) y las condiciones de exclusión (escoger o no los tumores basocelulares y los carcinomas *in situ*). Una vez ejecutada la aplicación, ésta diferencia los tumores prevalentes (previamente registrados en el RHT) de los incidentes (tumores nuevos para el centro sanitario) (fig. 1).

En el período 1999-2000 se detectaron 10.498 pacientes con algún código de malignidad en las bases

de datos de AH y APA de los centros sanitarios del ICO y la HUB. En la resolución manual realizada por el experto se detectaron en estos pacientes 10.683 tumores, de los cuales 8.309 (77,8%) correspondieron a tumores incidentes y 2.374 (22,2%) a tumores prevalentes (que ya se habían detectado en el RHT en el período 1990-1998) (fig. 2).

La aplicación ASEDAT resolvió 8.901 pacientes automáticamente (84,8%) quedando 1.597 pacientes para resolución manual (15,2%). En los 8.901 pacientes resueltos de manera automática se detectaron 9.017 tumores malignos. De éstos, 7.121 (79,0%) fueron tumores incidentes y el resto, 1.896 (21,0%), tumores prevalentes. En los 1.597 pacientes seleccionados para la resolución manual se detectaron 1.724 tumores. De éstos, 1.246 (72,3%) fueron tumores incidentes y 478 (27,7%) tumores prevalentes (fig. 3). La aplicación muestra por pantalla la información de los tumores seleccionados para la resolución manual y la de los tumores prevalentes como resultado de la resolución automática, para que pueda ser revisada por el experto (fig. 4). En el período de estudio, ASEDAT detectó 8.367 tumores incidentes, mientras que en la resolución manual se detectaron

**Figura 2. Resolución del RHT del ICO-HUB según un experto (procedimiento manual) y con la aplicación ASEDAT (procedimiento automatizado) del período 1999-2000.**



8.309 tumores. Por tanto, ASEDAT incrementó en un 0,7% (58 tumores) la detección de tumores incidentes.

La validación del grado de concordancia de la información de las variables del RHT se realizó con 7.063 tumores obtenidos a partir de los 7.121 tumores incidentes resueltos por el programa ASEDAT y excluyendo los 58 tumores que no fueron detectados de forma manual. De éstos, en 6.185 tumores (87,6%) la información coincidió en todas las variables (fecha de incidencia, localización tumoral, morfología, comportamiento y base diagnóstica). En el resto de tumores, en 878 (12,4%) hubo discordancia al menos en alguna de las variables consideradas. La variable que presentó mayor discordancia entre la resolución automática y la manual fue la morfología, en la que se detectaron 444 tumores discordantes (6,3%; índice kappa = 0,971), seguida de la base diagnóstica (n = 322; 4,6%; índice kappa = 0,917) y la localización tumoral (n = 279; 4,0%; índice kappa = 0,962) (tabla 2). En la variable fecha de incidencia se detectó discordancia en 108 casos (1,5%), y no se estimó el índice de kappa ya que se tuvo en cuenta la concordancia en 8 dígitos: día, mes y año.

De los 878 tumores discordantes, 692 (78,8%) fueron debidos a errores humanos generados en las fases de lectura, interpretación, selección, recodificación e introducción manual de la información de las bases de datos de APA y AH al RHT. La aplicación ASEDAT detectó este tipo de errores. En el resto de tumores, 186 (21,2%), los errores fueron consecuencia de la aplicación ASEDAT (fig. 4). Estas discordancias representan sobre el total de los casos ana-

lizados el 9,8 y el 2,6%, respectivamente. Los errores generados por la aplicación correspondieron a: a) error en la conversión de algunos códigos de la Universidad de Lovaina a la ICD-O (n = 63); b) errores de las bases de datos de APA y AH (n = 66); c) déficit de reglas específicas en algunos casos determinados (n = 49), y d) tumores hepáticos sin especificar en el AH si eran primarios o metastásicos (n = 8).

Para la automatización de la presentación de resultados del RHT se diseñó, en la página web del ICO, una página propia del RHT con el objetivo de que se realicen consultas abiertas y personalizadas *on-line* directamente a la base de datos del RHT (<http://rht.iconcologia.catsalut.net/cas/asedat.htm>). Para salvaguardar el anonimato de los pacientes se tuvieron en cuenta las normas de confidencialidad de datos.

La exportación de la aplicación ASEDAT al Hospital de la Cruz Roja de L'Hospitalet de Llobregat se realizó para el año 1999. La aplicación detectó 916 pacientes de los cuales 763 (83,3%) se resolvieron automáticamente y 153 (16,7%) pasaron a resolución manual. En los 763 pacientes automatizados se detectaron 774 tumores incidentes. La comparación entre la resolución automatizada y la manual de los tumores incidentes en las variables localización tumoral y morfología mostró concordancia en 672 tumores (86,82%) y diferencias en 102 tumores (13,18%). En cuanto a los tumores discordantes, 66 (64,7%) se debieron a errores humanos durante la resolución manual del registro y 36 (35,3%) a la aplicación ASEDAT. Estos errores representan el 8,53 y el 4,65% del total de casos, respectivamente.

Figura 3. Ejemplo de un caso seleccionado por la aplicación ASEDAT para ser revisado por un experto (revisión manual).

The screenshot shows the ASEDAT application window titled "ASEDAT - [Revisar un caso]". It contains two main tables:

**Entada AH y APA**

| HIC   | FONT | DATA_INC   | SERVEI | CODI_L | DESC_CODI_L   | CODI_M | DESC_CODI_M        |
|-------|------|------------|--------|--------|---|--------|--------------------|
| 81282 | A    | 17/07/2001 | CIR    | 153.3  | NEOPLASIA MALIGNA COLON SIGMOIDE; ANGLE SIGMOIDE          |        |                    |
| 81282 | A    | 17/07/2001 | CIR    | 041.11 | INFECCIO PER ESTAFILOCOCC AUREUS                          |        |                    |
| 81282 | A    | 17/07/2001 | CIR    | 491.20 | BRONQUITIS CRONICA OBSTRUCTIVA SENSE EXACERBACIO AGUDA    |        |                    |
| 81282 | A    | 17/07/2001 | CIR    | 518.5  | INSUFICIENCIA PULMONAR CONSECUTIVA A TRAUMATISME I CIRURG |        |                    |
| 81282 | A    | 17/07/2001 | CIR    | 303.91 | ALTRA I INESPECIFICA DEPENDENCIA DE L'ALCOHOL, CONTINUA   |        |                    |
| 81282 | A    | 17/07/2001 | CIR    | 934.9  | COS ESTRANY ARBRE RESPIRATORI LOC.NSP.INHALAC.LIQUID.VDMI |        |                    |
| 81282 | A    | 17/07/2001 | CIR    | V58.49 | ALTRA ASSISTENCIA ESPECIFICADA POSTERIOR A CIRURGIA       |        |                    |
| 81282 | A    | 17/07/2001 | CIR    | 573.8  | ALTRES TRASTORNS HEPATICS ESPECIFICATS; HEPATOPTOSI       |        |                    |
| 81282 | A    | 17/07/2001 | CIR    | 401.9  | HIPERTENSIO ESSENCIAL INESPECIFICADA                      |        |                    |
| 81282 | B    | 05/06/2001 | DIG    | T67700 | COLON SIGMOIDE  | M00100 | TEJIDO NORMAL      |
| 81282 | B    | 06/06/2001 | CIR    | T67000 | COLON, NE   | M81403 | ADENOCARCINOMA, NE |
| 81282 | B    | 06/06/2001 | CIR    | T55000 | GLANDULAS SALIVARES, NE                                   | M81403 | ADENOCARCINOMA, NE |

**Resolució manual**

| HIC   | Data       | Loc   | Modf   | DX | Servei | Prev | Font   | Revisat |
|-------|------------|-------|--------|----|--------|------|--------|---------|
| 81282 | 05/06/2001 | 153.3 | M81403 | 7  | CIR    |      | APA+AH | 1       |
| 81282 | 06/06/2001 | 142.9 | M81403 | 7  | CIR    |      | APA    | 1       |

Annotations in the image point to the following actions:

- Selección del resultado:** Points to the '+' button in the 'Resolució manual' table.
- Eliminación del caso seleccionado:** Points to the '-' button in the 'Resolució manual' table.
- Confirmación del caso seleccionado: automáticamente entrará en el RHT:** Points to the 'Revisat' column in the 'Resolució manual' table.

## Discusión

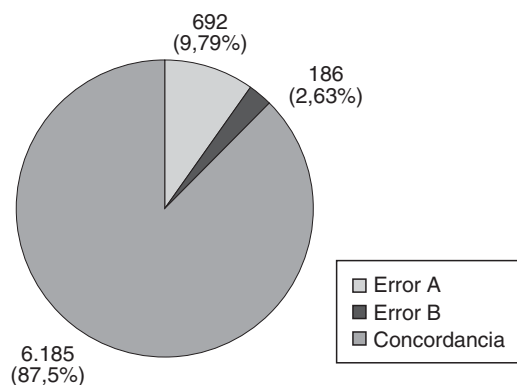
La automatización de un registro de cáncer hospitalario es posible siempre y cuando el centro disponga de las bases de datos de APA y AH informatizadas y codificadas con los códigos SNOMED e ICD-9, respectivamente. El porcentaje de resolución automatizada de los tumores incidentes obtenido en nuestro RHT mediante la aplicación ASEDAT ha sido del 84,8%. La automatización del RHT en nuestro centro ha presentado ventajas respecto a la resolución manual: *a)* ahorro en el coste de tiempo: el tiempo de resolución para los tumores incidentes se ha reducido a 2 h de cálculo computarizado frente a los 2 meses que se tardaba en recopilar la información de forma manual. El tiempo de resolución para los tumores prevalentes se redujo a 2 semanas frente al mes de la resolución manual; *b)* ahorro del coste económico: la selección, extracción, recodificación e introducción de la información automatizada en el 84,8% de los tumores incidentes ha reducido el tiempo de dedicación del personal experto y administrativo, y *c)* aumento de la fiabilidad de los resultados: reducción del error hu-

mano (9,8%) al generado por la propia aplicación (2,6%).

No hemos hallado experiencias de automatización en registros de cáncer hospitalarios, por lo que tenemos que compararnos con los resultados de las aplicaciones diseñadas para registros de cáncer poblacionales. Nuestra resolución ha sido superior a la observada en algunos registros de cáncer de base poblacional automatizados (el 51-69% de los registros de cáncer poblacionales)<sup>15,21,22</sup>. El mayor grado de resolución de nuestra aplicación es debido a que las fuentes de información de nuestro RHT son sólo 3 (AH, APA y el propio RHT). Por el contrario, las fuentes de información en los registros de cáncer poblacionales proceden de las bases de datos de AH y APA de todos los hospitales de un área geográfica determinada y de su respectivo registro de mortalidad. La diversidad de fuentes de información genera heterogeneidad de la información, mayor volumen de datos y la posibilidad de tener pacientes duplicados (pacientes asistidos en más de un hospital), disminuyendo el porcentaje de casos resueltos automáticamente<sup>12,13</sup>.

El grado de concordancia entre la resolución manual y automatizada con la aplicación ASEDAT del RHT

**Figura 4. Distribución de los tumores incidentes según la concordancia y el tipo de error tras la comparación de la resolución manual realizada por el experto y la resolución automatizada por ASEDAT.**



Error A: errores generados en la fase de lectura, interpretación, selección, recodificación e introducción de la información en la resolución manual.

Error B: errores generados por la aplicación ASEDAT en la resolución automática.

del ICO-HUB ha sido similar a la obtenida mediante las aplicaciones de los registros de cáncer de base poblacional (el 98,4% en el RHT del ICO-HUB y el 91,5-98,7% en los registros poblacionales)<sup>15</sup>. La similitud de la concordancia de las diferentes aplicaciones podría explicarse porque los tumores resueltos automáticamente presentan, en la mayoría de casos, un menor grado de complejidad y unas reglas de resolución mejor definidas<sup>12</sup>. Por tanto, los casos concordantes entre los 2 tipos de resolución son altos, con independencia del tipo de registro.

Una de las limitaciones de la automatización del RHT mediante ASEDAT, ajena a la propia aplicación, es que sólo recoge información de las bases de datos de APA y de AH. La informatización de las consultas externas en muchos hospitales no incluye los diagnósticos codificados para poder incluirse como una fuente de captación de pacientes, por lo que la extracción de la información de los pacientes ambulatorios sin recurrir a la revisión de la historia clínica es difícil. En nuestro hospital, hemos explorado la incorporación de nuevas fuentes de información para el RHT, como los servicios de farmacia (tratamientos con citostáticos, fecha de tratamiento, dosis, fármacos), oncología radioterápica (fecha y tipo de tratamiento con radioterapia) y quirófanos (fecha y tipo de cirugía). Una prueba piloto realizada para el año 2001 en la que se utilizaron estas bases de datos permitió recoger automáticamente la información de los tratamientos aplicados en los casos detectados mediante la aplicación ASEDAT, e incrementó en un 15% la detección de pacientes oncológicos, al incluir a pacien-

**Tabla 2. Diferencias en la información recogida en las variables básicas del registro hospitalario de tumores (RHT) entre los tumores detectados por el experto y los detectados por la aplicación ASEDAT**

| Variabes RHT                      | n                  | (%)   | Kappa | IC kappa    |
|-----------------------------------|--------------------|-------|-------|-------------|
| Fecha de la incidencia            | 108                | 1,53  |       |             |
| Localización tumoral <sup>a</sup> | 279                | 3,95  | 0,962 | 0,955-0,968 |
| Morfología <sup>a</sup>           | 444                | 6,29  | 0,971 | 0,966-0,976 |
| Método diagnóstico                | 322                | 4,56  | 0,917 | 0,908-0,927 |
| Tumores con alguna discordancia   | 878 <sup>b</sup>   | 12,43 |       |             |
| Tumores con total concordancia    | 6.185              | 87,60 |       |             |
| Total tumores incidentes          | 7.063 <sup>c</sup> |       |       |             |

IC kappa: intervalo de confianza del índice kappa.

<sup>a</sup>Concordancia en los 3 primeros dígitos del código topológico y de los 3 primeros dígitos numéricos del código morfológico. <sup>b</sup>Tumores que presentan discordancia entre la resolución manual y la automática en más de una variable. <sup>c</sup>Tumores incidentes del período 1999-2000 detectados en la resolución manual por parte del experto y en la aplicación ASEDAT.

tes ambulatorios (datos no expuestos). Estos resultados son similares a los hallados en otros estudios<sup>23</sup>, en los que se detectaba un mayor número de pacientes oncológicos en un centro al agregar información sobre los tratamientos aplicados.

Otra limitación importante es la imposibilidad de recoger de forma automatizada el estadio del paciente en el momento del diagnóstico. El estadio del tumor, aunque conste en los informes de AP o en la historia clínica, no está incluido en las bases de datos informatizadas de la mayoría de los centros sanitarios. Sería interesante que los hospitales realizaran un esfuerzo para incluir el estadio en el sistema informático, integrando el TNM patológico de los informes de AP y/o las decisiones de los comités de tumores.

La implementación de ASEDAT en el registro hospitalario del hospital de la Cruz Roja de L'Hospitalet de Llobregat confirmó que la aplicación puede ser exportable a otros centros. En Cataluña se está realizando un esfuerzo en el diseño de software para poder implementar un registro de cáncer poblacional de toda la comunidad autónoma. Las fuentes de información del registro poblacional serían: a) los registros de cáncer de base poblacional de las provincias de Tarragona y Girona; b) el registro de mortalidad de Cataluña, y c) para las provincias de Barcelona y Lleida, que no disponen de registros de cáncer, las bases de datos de AH y de APA de todos los hospitales (públicos y privados) y los registros de cáncer hospitalarios. Un programa informático de Record Linkage probabilístico<sup>24</sup>, diseñado conjuntamente con la Universitat Politècnica de Catalunya, identificaría a un mismo paciente que ha sido asistido en diferentes hospitales. La aplicación ASEDAT extraería, a partir de las bases de datos de AH, APA y de los registros hospitalarios de cada centro, la información de las variables esen-

ciales de cada tumor para el registro de cáncer de Cataluña.

En resumen, el diseño y la aplicación de sistemas informáticos facilita la identificación de pacientes diagnosticados de cáncer y la extracción de la información clínica del sistema sanitario para la constitución y el mantenimiento de un registro de cáncer (hospitalario o poblacional). Sin embargo, es necesario mejorar la calidad de la información de los centros sanitarios, informatizarla y no subestimar los recursos humanos destinados a supervisar, validar, confirmar y mejorar la información obtenida de forma automatizada.

### Agradecimientos

A la Dra. Carme Gimeno, del servicio de Documentación Médica del Hospital de la Cruz Roja de L'Hospitalet de Llobregat, por su inestimable ayuda en la implementación y validación de la aplicación ASEDAT en su centro, así como por sus comentarios que contribuyeron a mejorar procesos de la aplicación. Este trabajo ha sido financiado parcialmente por las Redes de Centros de Investigación Cooperativa en Cáncer del Instituto de Salud Carlos III (RTICCC C03/10) y por el Gobierno Catalán (DURSI 2003XT 00023).

### Bibliografía

- Borràs JM, Borràs J, Bosch FX, Fernández E, Galcerán J, Gispert R, et al. *Càncer Catalunya 2001*. L'Hospitalet de Llobregat: Institut Català d'Oncologia. Departament de Sanitat i Seguretat Social. Generalitat de Catalunya; 2001.
- Plan Director de Oncologia en Catalunya: 2001-2004. Barcelona: Institut Català d'Oncologia. Departament de Sanitat i Seguretat Social. Generalitat de Catalunya; 2001.
- Viladiu P. *Els registres hospitalaris de càncer*. Barcelona: Departament de Sanitat i Seguretat Social. Generalitat de Catalunya; 1993.
- Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB, editores. *Cancer incidence in five continents*. Vol. VIII. Lyon: IARC Scientific Publications. N.º 155.; 2002.
- Institut Català d'Oncologia. *Els nostres centres*. Doctor Josep Trueta. Girona. Unitat d'Epidemiologia i Registre de Càncer de Girona. Disponible en: <http://www.iconcologia.catsalut.net/doctor+josep+trueta/unitat+depidemiologia+i>
- Llistat de Codis Diagnòstics d'Anatomia Patològica segons les normes de la Universitat Catòlica de Lovaina. L'Hospitalet de Llobregat: Ciutat Sanitària i Universitària de Bellvitge; 2001.
- Cote RA, editor. *Systematized Nomenclature of Medicine (SNOMED)*. Vols I and II. 1st ed. Skokie: College of American Pathologists; 1977.
- Classificació Internacional de Malalties, 9.ª Revisió, Modificació clínica. 2.ª ed. Vols. I, II, III. Barcelona: Departament de Sanitat i Seguretat Social. Generalitat de Catalunya; 1993.
- Manual de Notificació del Conjunt Mínim Bàsic de Dades de l'Alta Hospitalària. Barcelona: Departament de Sanitat i Seguretat Social. Servei Català de la Salut. Generalitat de Catalunya; 1996.
- International Classification of Diseases for Oncology. 1st ed. Geneva: World Health Organization; 1976.
- Percy C, Van Holten V, Muir C, editores. *International Classification of Diseases for Oncology*. 2nd ed. Geneva: World Health Organization; 1990.
- Tyczynski JE, Démaret E, Parkin DM, editores. *Standards and guidelines for cancer registration in Europe. The ENCR recommendations*. Vol I. Lyon: IARC Technical Publication. N.º 40; 2003.
- Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, editores. *Cancer registration, principles and methods*. Lyon: IARC Scientific Publications; 1991.
- Ribes J, Clèries R, Sanz S, Díaz M, Muñoz J, Bosch FX. *Registre Hospitalari de Tumors '98*. L'Hospitalet de Llobregat: Servei d'Epidemiologia i Registre del Càncer. Institut Català d'Oncologia; 1998.
- Black RJ, Simonato L, Storm HH, Démaret E, editores. *Automated data collection in cancer registration*. Lyon: ARC Technical Reports; 1998.
- Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al, editors. *International Classification of Diseases for Oncology*. 3th ed. Geneva: World Health Organization; 2000.
- Gálvez J. *Automatització de un registre hospitalari de tumors*. Proyecto final de carrera. Facultat d'Informàtica de Barcelona. Barcelona: Universitat Politècnica de Catalunya; 2002.
- Sobin LH, Wittekind Ch, editores. *TNM Classification of Malignant Tumours*. 6th ed. New York: Wiley-Lyss, International Union Against Cancer (UICC); 2002.
- Methods in observational epidemiology. En: Kelsey JL, Whittemore AS, Evans AS, Thompson DW. *Methods in observational epidemiology*. Oxford: Oxford University Press; 1996.
- Institut Català d'Oncologia. *Informació General*. Epidemiologia, Prevenció i Control del Càncer. Registre Hospitalari de Tumors ICO/HUB. Disponible en: <http://www.iconcologia.catsalut.net/area+de+epidemiologia/registro+tumores.htm>
- Simonato L, Zambon P, Rodella S, Giordano R, Guzzinati S, Stocco C, et al. A computerised cancer registration network in the Veneto region north-east of Italy: a pilot study. *Br J Cancer*. 1996;73:1436-9.
- Middleton RJ, Gavin AT, Reid JS, O'Reilly DO. Accuracy of hospital discharge data for cancer registration and epidemiological research in Northern Ireland. *Cancer Causes Control*. 2000;11:899-905.
- Cooper GS, Yuan Z, Stange KC, Dennis LK, Amini SB, Rimm AA. Agreement of medicare claims and tumor registry data for assessment of cancer-related treatment. *Med Care*. 2000;38:411-21.
- Gómez-Bao J, Romera JR, Ribes J, Larriba-Pey JL. Fusión rápida y eficiente de múltiples ficheros con errores y falta de datos. Montpellier: XXIX Réunion du Groupe pour l'Epidemiologie et l'Enregistrement du Cancer dans les Pays de Langue Latine; 2004.